

1 Nitrogen cost minimization is promoted by structural changes in the transcriptome of N  
2 deprived *Prochlorococcus* cells

3

4 Robert W. Read<sup>1</sup>, Paul M. Berube<sup>2</sup>, Steven J. Biller<sup>2</sup>, Iva Neveux<sup>1</sup>, Andres Cubillos-  
5 Ruiz<sup>2,3</sup>, Sallie W. Chisholm<sup>2,4</sup>, and Joseph J. Grzymiski<sup>1\*</sup>

6

7 **1. Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, NV,**  
8 **USA**

9

10 **2. Department of Civil and Environmental Engineering, Massachusetts Institute of**  
11 **Technology, Cambridge, MA, USA**

12

13 **3. Microbiology Graduate Program, Massachusetts Institute of Technology,**  
14 **Cambridge, MA, USA**

15

16 **4. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA,**  
17 **USA**

18 Correspondence: Joseph J. Grzymiski, Department of Earth and Ecosystem Sciences,  
19 Desert Research Institute, 2215 Raggio Parkway, Reno, NV 89509, USA.

20 **\*E-mail: [joeg@dri.edu](mailto:joeg@dri.edu)**

21

22 **Conflict of interest.**

23 The authors declare no conflict of interest

## 24 **Abstract**

25 *Prochlorococcus* is a globally abundant marine cyanobacterium with many adaptations  
26 that reduce cellular nutrient requirements, facilitating growth in its nutrient-poor  
27 environment. One such genomic adaptation is the preferential utilization of amino acids  
28 containing fewer N-atoms, which minimizes cellular nitrogen requirements. We predicted  
29 that transcriptional regulation might be used to further reduce cellular N budgets during  
30 transient N limitation. To explore this, we compared transcription start sites (TSSs) in  
31 *Prochlorococcus* MED4 under N-deprived and N-replete conditions. Of 64 genes with  
32 primary and internal TSSs in both conditions, N-deprived cells initiated transcription  
33 downstream of primary TSSs more frequently than N-replete cells. Additionally, 117  
34 genes with only an internal TSS demonstrated increased internal transcription under N-  
35 deprivation. These shortened transcripts encode predicted proteins with ~5-20% less N  
36 content compared to full-length transcripts. We hypothesized that low translation rates,  
37 which afford greater control over protein abundances, would be beneficial to relatively  
38 slow-growing organisms like *Prochlorococcus*. Consistent with this idea, we found that  
39 *Prochlorococcus* exhibits greater usage of glycine-glycine motifs, which cause  
40 translational pausing, when compared to faster growing microbes. Our findings indicate  
41 that structural changes occur within the *Prochlorococcus* MED4 transcriptome during N-  
42 deprivation, potentially altering the size and structure of proteins expressed under nutrient  
43 limitation.

44

45

## 46 **Introduction**

47 Primary productivity is limited by nitrogen (N) availability in many ocean ecosystems  
48 (Tyrrell, 1999; Deutsch *et al.*, 2007; Moore *et al.*, 2013), and organisms that live there  
49 have adaptations that help them survive in low nutrient conditions. These adaptations  
50 include small cell sizes, which facilitate nutrient transport by increasing surface:volume  
51 ratios and reducing the absolute cellular requirement for nutrients (Munk and Riley,  
52 1952; Gavis, 1976; Chisholm, 1992), as well as small genomes and a proteome that is N  
53 cost minimized (Grzymiski and Dussaq, 2012). The concept of cost minimization was  
54 originally based on an observation that assimilatory proteins for a specific element  
55 contain relatively less of that specific element than average proteins in the same organism  
56 (Baudouin-Cornu *et al.*, 2001). For many oligotrophic microorganisms, however, cost  
57 minimization extends beyond the proteins involved in assimilation and is observed across  
58 the proteome; e.g. the genomes of N cost minimized oligotrophic microbes code for  
59 proteins that are, on average, reduced in amino acids containing added N side chains as  
60 compared to their coastal counterparts (Grzymiski and Dussaq, 2012). As an evolutionary  
61 trade-off, the proteomes of N-cost minimized organisms typically have slightly larger  
62 mass or more C atoms – an element which is not usually limiting for autotrophs  
63 (Grzymiski and Dussaq, 2012).

64

65 The marine cyanobacterium *Prochlorococcus* MED4 is a striking example of a N-cost  
66 minimized organism (Grzymiski and Dussaq, 2012). *Prochlorococcus* is often the  
67 numerically dominant primary producer in oligotrophic waters (Flombaum *et al.*, 2013),  
68 and can be broadly divided into two main subgroups, made up of high-light adapted and

69 low-light adapted cells. High-light adapted *Prochlorococcus* cells are the smallest  
70 cyanobacterial cells in the oceans, and also have the smallest genomes of any free-living  
71 photosynthetic cell. The abundance of high-light adapted *Prochlorococcus* cells typically  
72 exceeds that of low-light adapted cells by several orders of magnitude in surface waters  
73 (Ahlgren *et al.*, 2006; Zinser *et al.*, 2006; 2007; Malmstrom *et al.*, 2010). The divergence  
74 of high-light adapted *Prochlorococcus* from their low-light adapted relatives is correlated  
75 with a reduction in the overall G+C content of their genomes, thereby reducing the N  
76 requirements of these cells due to the bias of low G+C codons to encode low N-  
77 containing amino acids (Grzymiski and Dussaq, 2012).

78

79 In addition to reducing the N content of their proteomes, high-light adapted  
80 *Prochlorococcus* have undergone a process of genome streamlining in which the number  
81 of coding sequences has been reduced, in part, by dispensing with most regulatory  
82 proteins (Grzymiski and Dussaq, 2012; Giovannoni *et al.*, 2014). How, then, do these  
83 cells respond to changing environmental conditions? Recent studies have revealed that, in  
84 addition to traditional protein regulators (transcription factors, sigma factors, etc.), many  
85 bacteria also utilize RNA-based mechanisms to regulate transcript abundance (Sharma *et*  
86 *al.*, 2010). *Prochlorococcus*, like other streamlined microbes such as *Helicobacter pylori*,  
87 has an unexpectedly complex transcriptome for a small genome, that includes cis and  
88 trans regulating RNA, other non-coding RNAs, varying cistronic and polycistronic  
89 operon regulation, and mRNAs with short half-lives (Steglich *et al.*, 2010; Voigt *et al.*,  
90 2014). These regulatory characteristics likely enable responses to environmental signals  
91 without a complex protein-based regulatory network.

92  
93 Given the degree to which N cost minimization has been selected for in the genome of  
94 *Prochlorococcus* MED4 over evolutionary time, we wondered whether the transcriptional  
95 response in these cells has also been optimized to further reduce N requirements. We  
96 hypothesized that transcriptome changes could help provide additional N savings during  
97 nutrient limited growth – a concept we define as transcriptomic cost minimization. To  
98 test this hypothesis, we compared the primary transcriptomes (the set of unprocessed  
99 transcripts that allow experimental determination of transcription initiation sites) of  
100 *Prochlorococcus* MED4 grown under N-deprived and N-replete conditions, allowing us  
101 to examine alterations in the transcriptional start sites under N-deprivation. Our results  
102 uncovered responses to N stress that highlight the capacity of already ‘streamlined’ cells  
103 to dynamically minimize their N requirements in response to N-deprivation.

104

## 105 **Materials and Methods**

### 106 **Cell cultures**

107 Axenic *Prochlorococcus* MED4 cells were grown in batch culture at 25 °C under  
108 continuous illumination of  $\sim 55 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$  using cool, white, fluorescent bulbs.  
109 Cultures were grown in Pro99 media (Moore *et al.*, 2007), prepared with 0.2  $\mu\text{m}$  filtered  
110 surface seawater from the South Pacific Subtropical Gyre (26.25 °S, 104 °W) and  
111 amended with 3.75 mM TAPS buffer (pH=8) and 6 mM sodium bicarbonate to control  
112 pH. Cells were allowed to acclimate to these temperature, light and media conditions by  
113 successively transferring the cultures four times during mid-exponential phase ( $\sim 20$   
114 generations).

115

116 Triplicate 4.5 L batch cultures were grown in 10 L clear polycarbonate carboys. Bulk  
117 culture fluorescence was measured daily using a 10AU fluorometer (Turner Designs,  
118 Sunnyvale, CA, USA). Cells were preserved for flow cytometry by fixation in 0.125%  
119 glutaraldehyde and storage at -80 °C. Once the cultures reached mid-exponential phase,  
120 3.8 L of each culture was concentrated by centrifugation in four 1 L centrifuge bottles  
121 (Beckman Coulter, Brea, CA, USA) using a JLA-8.1000 rotor in an Avanti J-20 XP  
122 centrifuge (Beckman Coulter, Brea, CA, USA) at 25 °C and 9000 g (6857 rpm) for 15  
123 min. Each cell pellet was washed once by re-suspension in 375 mL of N-depleted media  
124 (lacking ammonium chloride addition) and concentrated again by centrifugation. Half of  
125 the washed cells were re-suspended in 1.8 L of N-replete Pro99 media, and the remaining  
126 half of the washed cells were re-suspended in 1.8 L of N-depleted Pro99 media (lacking  
127 ammonium chloride addition).

128

129 Samples for bulk culture fluorescence, cell enumeration by flow cytometry, fluorescence  
130 induction measurements, and RNA extraction were collected at 0, 3, 12 and 24 hours  
131 following the re-suspension of cells in either nitrogen-replete or nitrogen-depleted media.  
132 At each time point, bulk culture fluorescence was measured using a 10AU fluorometer  
133 (Turner Designs, Sunnyvale, CA, USA), and cells were preserved for flow cytometry by  
134 fixation in 0.125% glutaraldehyde and storage at -80 °C. Following the experiment, fixed  
135 cells were enumerated using an Influx Cell Sorter (BD Biosciences, San Jose CA, USA),  
136 as previously described (Olson *et al.*, 1985; Cavender-Bares *et al.*, 1999). Photochemical  
137 conversion efficiency ( $F_v/F_m$ ) was measured using a FIRE Fluorescence Induction and

138 Relaxation System (Satlantic, Halifax, Nova Scotia, Canada); cells were acclimated in the  
139 dark for 30 min to relax photosynthetic reaction centers before fluorescence induction  
140 curves were obtained using single turnover flash (STF) with blue light (450 nm with 30  
141 nm bandwidth). Raw data were processed using fireworx (Dr. Audrey B. Barnett,  
142 Dalhousie University, Halifax, Nova Scotia, Canada) in MATLAB. Specific growth rates  
143 of replicates in the 4.5 L batch starter cultures were estimated from the log-linear portion  
144 of growth curves constructed from bulk culture fluorescence and cell counts obtained  
145 from flow cytometry analysis. Fluorescence-based measurements yielded a growth rate of  
146  $0.60 \text{ d}^{-1} \pm 0.01$  (1 standard deviation) and flow cytometry based measurements yielded a  
147 growth rate of  $0.67 \text{ d}^{-1} \pm 0.14$  (1 standard deviation).

148

149 At each time point, biomass for RNA extraction was obtained by centrifugation of 250  
150 mL of culture in 250 mL centrifuge bottles (Beckman Coulter, Brea, CA, USA) using a  
151 JA-14 rotor in an Avanti J-20 XP centrifuge (Beckman Coulter, Brea, CA, USA) at 25 °C  
152 and 9000 g (9666 rpm) for 15 min. Cell pellets were re-suspended in 750  $\mu\text{l}$  of RNA  
153 Denaturation Solution (Ambion ToTALLY RNA Total RNA Isolation Kit, Life  
154 Technologies, Grand Island, NY, USA) by pipetting, frozen in liquid nitrogen and stored  
155 at -80 °C until processing, at which point total RNA was extracted using the Ambion  
156 ToTALLY RNA kit (Life Technologies, Grand Island, NY, USA), yielding ~12  $\mu\text{g}$  of  
157 total RNA from each pellet.

158

159 **RNA sequencing overview**

160 Total RNA from each sample was divided and used in two sequencing workflows: gene  
161 expression analysis using strand specific sequencing of terminator exonuclease (TEX)-  
162 treated RNA (all time points) and primary transcriptome analysis using 5' rapid  
163 amplification of cDNA ends (RACE) tagRNA-seq (12 and 24 hour time points). For gene  
164 expression analysis, it is necessary to account for transcripts within the total RNA pool  
165 that have undergone triphosphate to monophosphate conversion for ultimate degradation  
166 (Celesnik *et al.*, 2007; Schoenberg, 2007). A caveat of doing time-series analysis of gene  
167 expression using total RNA is that these signals are not removed and changes can be  
168 obfuscated, especially if half-lives of RNA are pathway or gene-dependent – a likely  
169 scenario with *Prochlorococcus*. The half-life differences of RNAs (Steglich *et al.*, 2010)  
170 could play a significant role in interpreting differences in RNA treatment methodology.  
171 For these reasons, we chose to use TEX-treated RNA for gene expression analysis. For  
172 primary transcriptome analysis and determination of transcription start sites, 5' RACE  
173 tagRNA-seq was used because it selectively captures 5' triphosphate ends at single  
174 nucleotide resolution. This produces the sawtooth-like pattern observed at transcriptional  
175 start sites throughout the genome (Sharma *et al.*, 2010).

176

### 177 **Ion Torrent sequencing for gene expression analysis**

178 RNA was treated with Terminator Exonuclease (Epicentre, Madison, WI, USA) to  
179 remove rRNA contamination and sequenced at the Nevada Genomics Center (Reno, NV,  
180 USA). Library preparation was performed using the Ion Total RNA-Seq Kit v2 (Life  
181 Technologies, Grand Island, NY, USA), which fragmented the total RNA and used  
182 reverse transcriptase to produce cDNA for strand-specific sequencing on the Ion



183 platform. Strand specific Ion Torrent Sequencing (Life Technologies, Grand Island, NY,  
184 USA) yielded 130 bp reads. Raw sequencing reads were uploaded to the sequencing read  
185 archive (NCBI accession: SRP078366, SRX1939126), and were inspected using FastQC  
186 (Andrews, 2009) to determine quality, ambiguous read percentage and relative amount of  
187 sequence reads. Aliases for the individual run names can be found at  
188 <https://trace.ddbj.nig.ac.jp/DRAsearch/experiment?acc=SRX1066875>. Ion-Torrent RNA  
189 sequencing resulted in an average of  $2.62 \times 10^7 \pm 5.77 \times 10^6$  raw reads per cDNA library for  
190 the terminator exonuclease (TEX)-treated, N-deprived cells and  $2.29 \times 10^7 \pm 6.99 \times 10^6$   
191 reads for the TEX-treated, N-replete cells. The average quality score was Q24  
192 (Supplementary Table S1).

193

194 Sequencing reads were utilized as the input for Rockhopper (McClure *et al.*, 2013) for  
195 bacterial RNA-sequencing analysis. Seed length was 15% of the read length, and any  
196 read with mismatches greater than 15% of the read was disregarded. This removed  
197 approximately 10% of the reads from each treatment. Differential expression was  
198 calculated against a negative binomial distribution between the experimental (N-  
199 deprived) cells and the control (N-replete cells) at each time point (0, 3, 12, 24 hours post  
200 N removal) using the DESeq algorithm (Anders and Huber, 2010; Tjaden, 2015).

201 Significantly differentially abundant transcripts were determined as those with a *p*-value  
202  $< 0.05$  when comparing the two treatments. Pearson correlation coefficients to previous  
203 research (Tolonen *et al.*, 2006) were larger for genes in the top 50% of mean expression  
204 level; these differences are likely due to low read depth mapping in non-regulated genes.

205 Therefore, expression changes are only reported for those transcripts with normalized  
206 Rockhopper expression values in the top 50% of the data set.

207

### 208 **Illumina sequencing for primary transcriptome analysis**

209 To determine the primary transcriptome of MED4, we analyzed transcriptional mappings  
210 at 12 and 24 hours post N-deprivation relative to the N-replete controls. Total RNA was  
211 treated with Ribo-Zero rRNA Removal Kit for bacteria (Epicentre) to remove rRNA and  
212 any processed transcripts with 5' monophosphate ends. The samples were next treated  
213 with 5' RNA polyphosphatase (5RPP; Epicentre) in order to convert the remaining 5'  
214 triphosphate structures into 5' monophosphate ends in preparation for adapter ligation. 5'  
215 Illumina TruSeq adapters were ligated to the monophosphate groups of the transcripts  
216 and cDNA was synthesized. The tagged 5' cDNA fragments were then specifically  
217 amplified with PCR and sequenced using 5' RACE tagRNA-seq on a NextSeq 500  
218 system with 75 bp read lengths (vertis Biotechnologie AG). To visualize the sequencing  
219 reads (NCBI accession: SRP078366, SRX1939254) and transcriptional changes, reads  
220 were mapped to the *Prochlorococcus* MED4 genome using the segemehl short read  
221 aligner (Hoffmann *et al.*, 2009; 2014). Resulting sam files containing mapped reads were  
222 converted into sorted and indexed bam files using samtools (Li *et al.*, 2009). A single  
223 alignment run of Transcription Start Site Annotation Regime (TSSAR) software was  
224 performed to determine start site differences between the transcriptional start site  
225 enriched samples and non-enriched samples in the N-replete and N-depleted samples  
226 (Amman *et al.*, 2014). In keeping with the definitions considered by the TSSAR  
227 algorithm, we classified TSSs located on the opposite strand of an annotated gene as

228 antisense; TSSs located within 250 nt upstream of the gene's annotated TSS as primary  
229 start sites; and TSSs within the annotated gene as internal (Amman *et al.*, 2014). Orphan  
230 TSSs were those that did not fall into any of these three categories, and we noted that  
231 some TSSs can represent a combination of multiple categories. Transcription from  
232 internal start sites in the sense orientation yield mRNAs that are called intraRNAs.  
233 Parameters for determining start site changes were a p-value of  $1 \times 10^{-10}$ , a noise threshold  
234 of 4 and a merge range of 5. All TSSs identified were hand-curated by uploading the  
235 sorted bam files into the Integrative Genomics Viewer, which placed the reads in the  
236 correct position and orientation in relation to the annotated *Prochlorococcus* MED4  
237 genome (Robinson *et al.*, 2011; Thorvaldsson *et al.*, 2013). Positions with a read count  
238 difference of less than 100 between the enriched and non-enriched samples were ignored.  
239 The TSSs identified by TSSAR after 24-hour post N-depletion were used to determine  
240 which TSSs were conserved between 12 and 24 hours post N-depletion. Sorted bam files  
241 were opened in R and read coverage figures were produced using the ggplot2 package  
242 (Wickham, 2009). If both a primary and internal TSSs were present, the abundance of  
243 transcripts derived from an internal TSS was calculated by comparing the read mapping  
244 at the internal TSS to the read mapping at the primary TSS for both the N-depleted and  
245 N-replete samples. The magnitude of internal transcription in each treatment was then  
246 compared against each other. Only internal TSSs resulting in transcript abundance greater  
247 than 10% compared to their primary TSS were used. If no primary TSS was present,  
248 transcription from internal TSS was calculated by directly comparing read mapping in the  
249 N-depleted sample to read mapping in the N-replete sample.  
250

## 251 **Multiple sequence alignment and protein threading**

252 For several genes with possible NtcA binding sites, we compared the structure of  
253 isoforms corresponding to those derived from transcripts initiated from internal start sites  
254 with PDB proteins using the protein structure predicting algorithm Phyre2 (Mezulis *et al.*,  
255 2015). Corresponding structure predictions were then overlaid on top of each other using  
256 separate colors to highlight differences between the two structures. Full-length proteins  
257 were also uploaded to NCBI and a domain search was completed using the conserved  
258 domain database (CDD). A caveat of this method is that no kinetics can be established,  
259 thus the functional efficiency of these predicted proteins remains unknown.

260

## 261 **Motif Analysis**

262 Motif analysis of the Gly-Gly motifs and analysis of the 12mer motifs for pyrimidine and  
263 purine frequencies were carried out according to published methods (Grzymiski *et al.*,  
264 2014). Gly-Gly motifs were calculated from coding regions broken up by triplets to  
265 account for codon frequencies.

266

## 267 **Results and Discussion**

### 268 **Physiological responses to N-deprivation**

269 Our experimental design was based on previous work examining abrupt N deprivation in  
270 *Prochlorococcus* MED4 (Tolonen *et al.*, 2006), facilitating comparisons between  
271 previous microarray gene expression results and our analysis of the primary  
272 transcriptome. Exponentially growing cells were subjected to acute N stress by  
273 resuspension in media with no added N (N-depleted). N-replete media was used as a

274 control. Bulk culture fluorescence values began to diverge three hours after resuspension  
275 in N-depleted media; the fluorescence of N-replete cultures continued to increase while  
276 the fluorescence of N-depleted cultures decreased (Fig. 1). Maximum PSII photochemical  
277 efficiency, Fv/Fm, was used as a measure of the cells' physiological response to N  
278 deprivation (Parkhill *et al.*, 2001). Fv/Fm values for both N-replete and N-deprived cells  
279 declined during the first three hours, likely due to the shock of centrifugation and  
280 resuspension, but stabilized in the N-replete cells within 12 hours. By contrast, Fv/Fm  
281 continued to decline in the N-deprived cells during the course of the experiment (Fig. 1).  
282 Following 24 hours in the N-depleted media, the N-deprived cultures had lower bulk  
283 fluorescence, fewer cell counts and lower Fv/Fm values compared to N-replete cultures.  
284 These results are consistent with previous work (Tolonen *et al.*, 2006) and indicate a  
285 significant physiological response of *Prochlorococcus* MED4 to N-deprivation.

286

### 287 **MED4 exhibits a canonical gene expression response to N-deprivation**

288 Using the Ion Torrent reads derived from TEX-treated RNA, we first examined overall  
289 gene expression patterns by comparing the transcriptomes of the N-deprived and N-  
290 replete cells at 3 different time points following induction of N stress. Overall, the gene  
291 expression changes we observed in the primary transcriptome in response to N  
292 deprivation were similar to changes previously observed in microarray analysis of total  
293 RNA (Tolonen *et al.*, 2006) (Supplementary Text and Supplementary Table S2). We  
294 found that the relative changes of differentially expressed transcripts, between N-  
295 deprived and N-replete cells, were approximately symmetrical; the number of transcripts  
296 demonstrating an increased abundance was similar to the number of transcripts exhibiting

297 a decreased abundance. These data indicate a balanced expression response to N  
298 deprivation rather than decreasing transcription overall (Table 1). Down-regulation of a  
299 substantial fraction of genes would be a rapid N-saving mechanism and was not the case  
300 in *Prochlorococcus* MED4, indicating an active response to N deprivation compared to  
301 entry into stasis. Further, the transcriptional response was rapid, with significant changes  
302 in transcript abundance apparent within 3 hours of N-deprivation and additional changes  
303 noted after 12 and 24 hours (Supplementary Tables S3-S6).

304

305 Cyanobacteria ‘sense’ N-deprivation by monitoring the C:N balance of the cell and  
306 respond to changes in this ratio by activating alternative N assimilation pathways when  
307 excess intracellular 2-oxoglutarate accumulates (Ohashi *et al.*, 2011). The global nitrogen  
308 stress response regulator, NtcA (Sauer *et al.*, 1999), is responsible for inducing the  
309 expression of assimilation pathways for a variety of N compounds. Consistent with  
310 previous reports (Tolonen *et al.*, 2006), we found that several members of the NtcA  
311 regulon are induced in response to N deprivation in *Prochlorococcus* MED4  
312 (Supplementary Table S5). Members of the NtcA regulon that exhibited increased  
313 transcript abundance during N-deprivation included genes encoding the assimilation  
314 pathways for urea and cyanate, as well as the glutamine synthase gene *glnA*  
315 (Supplementary Tables S3-S6). The NtcA regulon also regulates the urease enzyme and  
316 urea transporter enzyme. All subunits of the urease enzyme (*ureA-C*) were co-expressed  
317 and showed similar increases in relative abundance at 12 hours post N deprivation. The  
318 urea transporter genes (*urtA-E*) were transcribed in a similar manner with the reads  
319 gathering at the TSS for *urtA*, the first gene in the operon (Fig. S1). In addition, several

320 high light inducible proteins, known to respond to stress conditions in *Prochlorococcus*  
321 (Havaux *et al.*, 2003; Tolonen *et al.*, 2006; Steglich *et al.*, 2008), increased in abundance  
322 after 12 hours of N deprivation (Supplementary Table S7). One of these high light  
323 inducible proteins, Hli10, is encoded by a gene predicted to be regulated by NtcA  
324 (Tolonen *et al.*, 2006).

325

### 326 **Complexity of transcription initiation at variable start sites under N-deprivation**

327 Based on the Ion Torrent data set, gene expression changes in these *Prochlorococcus*  
328 MED4 cells exhibited responses similar to previous studies under N-deprivation (Tolonen  
329 *et al.*, 2006; Gilbert and Fagan, 2011) suggesting that N savings cannot be achieved by  
330 simply downregulating transcription. Thus, to investigate whether *Prochlorococcus*  
331 might use other RNA-based regulatory mechanisms to reduce cellular N requirements  
332 under N stress, we next examined the primary transcriptome of MED4 following 12 and  
333 24 hours of N depletion. We performed detailed annotation of TSSs after 12 hours post  
334 N-depletion due to the strong correlation of transcript patterns at this time point as  
335 compared to Tolonen *et al.* (2006), and confirmed these results with data from 24 hours  
336 post N-depletion (Table 2, Supplementary Tables S8-S11).

337

338 Our detailed annotation of TSSs at 12 and 24 hours following N-depletion identified new  
339 internal and antisense TSSs (Table 2, Supplementary Tables S8-S11) throughout the  
340 genome that were not previously reported by Voigt *et al.* (2014) for MED4 grown under  
341 N-replete conditions (Supplemental Text, Supplementary Table S12). For the purpose of  
342 this study, internal transcription is defined as a verified a saw-tooth pattern of transcript

343 reads mapping downstream of a gene's annotated start site. The internal transcription  
344 ratio is defined as the ratio of read abundance at the internal TSS compared to the  
345 abundance of reads at the primary TSS. The internal transcription ratio generally  
346 increased under N-deprivation (Fig. 2; Supplementary Table S13), suggesting a specific  
347 regulatory response to N stress rather than erroneous transcription given that we did not  
348 observe degraded and random read mapping to these genes. Eighty-five genes contained  
349 both a primary TSS and internal TSS identified by TSSAR in both the N-depleted and N-  
350 replete samples (Supplementary Table S13). Based on a Wilcoxon rank sum test, 64  
351 genes (~74%) demonstrated a significant increase in the internal transcription ratio (p-  
352 value  $< 1 \times 10^{-6}$  for all 64 genes) in N-depleted samples compared to the N-replete  
353 samples, with 13 genes containing multiple internal TSSs (for a total of 81 TSSs in these  
354 genes). After 24 hours of N deprivation, transcripts expressed from 72/81 (89%) of these  
355 internal TSSs (from 56 genes) were still present, indicating a conserved and continued  
356 role in the N stress response.

357

358 Transcription from an internal start site, however, was not exclusively a response to  
359 nutrient deprivation. We found the internal transcription ratio increased in 13 of the 85  
360 genes (15%) during N-replete conditions compared to N-deprivation, with an additional 8  
361 genes (11%) exhibiting a similar internal transcription ratio in both the N-replete and N-  
362 depleted samples (Supplementary Table S13). Irrespective of condition (N-deplete vs. N-  
363 replete) 35 genes (~41%) showed a higher abundance of internal TSS reads compared to  
364 primary TSS reads, indicating a general increase in internal transcription for those genes.  
365 Combined with 187 TSSs from 160 genes that contain only an internal TSS, of which 117



366 (63%) demonstrated greater internal transcription in N-depleted cells, these results could  
367 indicate that some internal TSSs are actually the primary transcription start site which  
368 may have been incorrectly called. Consistent with this, Sharma et al. found 18 new gene  
369 start sites in their re-annotation of *Helicobacter pylori* using a similar method (Sharma *et*  
370 *al.*, 2010).

371

372 On the other hand, we also observed that many of these sites are located far downstream  
373 of the annotated start site and well within the expected coding sequence. This suggests  
374 that some of these internal TSSs may yield short sense RNAs (intraRNAs) - molecules  
375 which have been identified in other studies and are hypothesized to function as truncated  
376 alternative mRNAs (Mitschke *et al.*, 2011; Shao *et al.*, 2014). Antisense TSSs were  
377 abundant in both N-replete and N-deprived cells, however the relative number of them  
378 found throughout the genome was approximately equal in both treatments and cannot be  
379 correlated to gene expression changes. These findings further highlight the incredible  
380 complexity of transcriptional regulation in *Prochlorococcus*.

381

### 382 **Transcriptional regulation and the N-cost minimization of proteins**

383 Many of the most abundant proteins found within microbes in the Sargasso Sea are  
384 transport proteins (Sowell *et al.*, 2008), emphasizing their importance to bacteria living in  
385 extremely nutrient depleted environments. NtcA regulation of transporters is known to be  
386 an important part of the response of *Prochlorococcus* cells to N-deprivation, as evidenced  
387 by a marked increase in transcripts encoding the urea transporter Urt and ammonium  
388 transporter Amt1 under these conditions (Tolonen *et al.*, 2006). Based on this, we next

389 focused on the regulation of other genes with possible NtcA binding sites, which could  
390 allow them to be regulated as part of the NtcA regulon.

391

392 As described above, N-deprivation increased internal transcription in *Prochlorococcus*  
393 MED4 (Fig. 2). The intraRNAs initiated from these internal TSSs could have three  
394 possible functions: 1) they could be mis-transcribed, degraded or processed RNA with  
395 unknown or no function; 2) they could be structural RNAs or encode peptide scaffolds  
396 that are used for other regulatory purposes (Lybecker *et al.*, 2014; Shao *et al.*, 2014); or  
397 3) they could be internally transcribed genes that code for functional proteins. While the  
398 intraRNAs could arise from transcriptional noise due to spurious promoter-like  
399 sequences, and in that case would be non-functional copies of RNA (Lybecker *et al.*,  
400 2014; Shao *et al.*, 2014; Thomason *et al.*, 2015), several genes with internal TSSs have  
401 clear regulatory differences in response to N-deprivation (Fig. 2). This has been seen in  
402 other cyanobacteria; for example, transcriptional mapping in *Synechocystis* sp. PCC6803  
403 identified abundant short sense transcripts from internal TSSs, which were proposed to  
404 yield shorter isoforms of *Synechocystis* 6803 proteins (Mitschke *et al.*, 2011). Random  
405 processes would not be expected to show reproducible responses to N-deprivation (i.e.,  
406 an indication of transcription level regulation). Thus, it is likely these transcripts have a  
407 specific physiological role.

408

409 We observed that many genes with internal transcription under N-deprived conditions  
410 were responsible for important physiological functions. This fact led us to hypothesize  
411 that these intraRNAs may encode functional proteins; the shortened RNA transcripts and

412 shortened translated proteins would reduce the overall N requirements of the cell. Given  
413 that these cells were already nutrient stressed and exhibiting significantly reduced  
414 photosynthesis and growth rates (Fig. 1), the cell might derive a net fitness benefit by  
415 expressing a partial protein which required fewer nutrients, as long as that protein  
416 retained at least some functionality. To examine this hypothesis, we evaluated the  
417 potential effect of the N-terminal truncation on the predicted functional domains of the  
418 proteins using NCBI's conserved domain search (Supplementary Table S14). We also  
419 used multiple sequence alignments (MSA) and protein threading to predict the structure  
420 of multiple proteins translated from intraRNAs by comparison to known structures  
421 (Supplementary Table S14). All of the proteins we examined by protein threading  
422 (Supplementary Table S14) contain a possible NtcA binding site upstream of the internal  
423 TSS, providing a potential mechanism for internal transcription initiation. We predict that  
424 initiating translation at internal start sites would likely have one of three outcomes: 1)  
425 proteins that align by MSA to all of the major structural elements of known structures  
426 and retain the major carbon backbone; 2) proteins that lack all structural elements or  
427 whose structure is clearly incomplete and; 3) proteins where a few specific structural  
428 elements or domains are transcribed (Supplementary Table S14).

429

430 The cyanate ABC transporter (PMM0370) evidences no structural differences between  
431 the full length translated protein and the shortened protein (Fig. 3a, Supplementary Table  
432 S14). Although the full diversity of functional cyanate transport proteins with established  
433 Protein Data Bank PDB structures have yet to be discovered, our conserved domain  
434 analysis results suggest that the shortened protein retains full function of its two predicted

435 domains (Supplementary Table S14). Still, the combined primary transcriptome and  
436 modeling results highlight how little we know about many bacterial proteins and further  
437 biochemical analysis are warranted to explore how changes in the primary amino acid  
438 sequence impact protein function.

439

440 Additionally, a number of other non-transport proteins may have been expressed in  
441 shorter forms. For ferredoxin-NADP reductase, there are no discernible changes in the  
442 threaded protein structure of the internal transcript, and it likely retains full function of its  
443 single predicted domain (Fig. 3b, Supplementary Table S14). RNA polymerase, vital for  
444 producing actual transcripts, has an internal TSS whose transcript encodes a protein that  
445 aligns to 95% of the sequence from the known PDB structure from *E. coli*, although  
446 domain function could be slightly modified as the entire protein is predicted as a single  
447 domain (Fig. 3c, Supplementary Table S14). It makes sense, physiologically, that under  
448 N-stress conditions the protein responsible for producing transcripts is potentially N cost-  
449 minimized. If these three proteins were internally translated they would provide 541 mol  
450 N / mol translated protein savings relative to the fully translated proteins – approximately  
451 17% savings. The three internal transcripts would provide 4451 mol N/ mol transcript  
452 savings compared to the full-length transcripts – a savings of, again, 17%.

453

454 In the context of the cell as a whole, transcriptomic N cost minimization is but one of  
455 several mechanisms through which *Prochlorococcus* cells adjust their elemental  
456 requirements in response to particular stressors such as limitation for key elements such  
457 as iron, phosphorus, and nitrogen. For example, in response to iron limitation

458 *Prochlorococcus* expresses the non-iron containing oxidoreductase flavodoxin instead of  
459 the iron-sulfur containing ferredoxin (Bibby *et al.*, 2003; Thompson *et al.*, 2011).  
460 *Prochlorococcus* also utilizes sulfolipids instead of phospholipids in its cellular  
461 membrane in order to decrease phosphorus requirements (Van Mooy *et al.*, 2006). These  
462 responses are all based on the utilization of different pathways in order to modify the  
463 cell's elemental requirements. In contrast, transcriptomic cost minimization represents a  
464 response to nutrient limitation that depends on structural changes to the mRNA pool and,  
465 putatively, the proteome. While genomic N cost minimization, mediated by codon usage  
466 and general %GC characteristics, can only change over evolutionary time scales,  
467 transcriptomic N cost minimization is a dynamic process which enables the cell to  
468 respond to changes in N availability on the order of hours.

469

#### 470 **Cost minimized organisms have a higher potential for decreased translation rates**

471 An important consideration regarding both genomic and transcriptomic N cost  
472 minimization is that such N savings might only have a small overall impact if protein  
473 abundance is not tightly controlled. One strategy for maintaining careful control of  
474 protein levels is rapid RNA turnover (Steglich *et al.*, 2010), which allows the cell to  
475 quickly adjust mRNA availability in response to stress. *Prochlorococcus*, for example,  
476 has a median RNA half-life on the order of 2 minutes, which is twofold faster than that  
477 observed in other microorganisms (Steglich *et al.*, 2010). Cells can also improve their  
478 control of protein abundance by slowing down translation rates (Sherman and Qian,  
479 2013). While faster growing (r-selected) organisms should experience selective pressures  
480 to maximize translation rates in order to support episodes of rapid growth, we

481 hypothesize that slowly growing (k-selected), cost minimized organisms should instead  
482 experience selective pressures to minimize translation rates. To explore this hypothesis,  
483 we examined the genomes of both oligotrophic k-strategists and copiotrophic r-strategists  
484 to look for sequence signatures associated with translation rates.

485

486 Shine-Dalgarno motifs (ribosomal binding sites) internal to coding regions have been  
487 shown to cause translational pausing and subsequent reductions in growth rate (Li *et al.*,  
488 2012). These signatures are apparent in the observed-to-expected ratio of the frequency  
489 of the glycine-glycine motifs in coding regions. Depending on codon usage, Gly-Gly  
490 motifs can have high or low affinities to the anti-Shine-Dalgarno sequence found at the 3'  
491 terminus of the 16S rRNA in the ribosome to mediate pausing; such sequences are  
492 minimized in fast-growing organisms like *E. coli* (Li *et al.*, 2012). We examined the  
493 occurrence of Gly-Gly motifs in the genomes of three bacteria with relatively high  
494 maximum growth rates (*Bacillus cereus*, *E. coli* K12 and *Vibrio fisheri*) and four  
495 oligotrophic, k-selected marine bacteria with relatively slow maximum growth rates  
496 (*Prochlorococcus* MED4, *Prochlorococcus* MIT9313, *Synechococcus* and *Pelagibacter*)  
497 (Fig. 4). We found that the genomes of the r-selected organisms exhibited a clear pattern  
498 of minimization of Gly-Gly motifs with high affinity to the anti-Shine-Dalgarno  
499 sequence, consistent with less pausing and faster protein expression. In oligotrophic k-  
500 selected organisms, the pattern is opposite with no apparent deviations in the observed-  
501 expected ratios of the motifs; in fact, the motifs with the highest binding affinity to the  
502 anti-Shine-Dalgarno sequence were found more often than expected (observed:expected  
503 = 1.63) in *Prochlorococcus* MED4. This suggests that selection against motifs that cause

504 translational pausing is weak in oligotrophic k-selected organisms, and therefore that they  
505 would be expected to have slower translation rates than the r-strategists. This genomic  
506 feature represents yet another property that may contribute to cost minimization in  
507 oligotrophic microbes.

508

509 We propose that there are at least two mechanisms contributing to this weak selection  
510 against pausing sequences in slow-growing, oligotrophic bacteria. First, as previously  
511 discussed, translational pausing would be beneficial to k-selected organisms by  
512 improving the ability of cells to control protein abundance while growing in nutrient  
513 limited conditions. Second, it is likely that organisms such as *Prochlorococcus* do not  
514 rely on Shine-Dalgarno sequences for translation initiation (Voigt *et al.*, 2014). Instead,  
515 based on studies in *Synechococcus*, translation initiation may instead rely on an  
516 alternative mechanisms dependent on ribosomal protein S1 (Mutsuda and Sugiura, 2006;  
517 Voigt *et al.*, 2014) or the direct binding of a 70S monosome to leaderless mRNA start  
518 sites (Moll *et al.*, 2002). This would reduce selection for canonical ribosomal binding  
519 sites and, as a less specific mechanism for translation initiation, might provide cells with  
520 a way to translate truncated proteins from mRNAs expressed from internal start sites  
521 under N stress. These k-selected organisms utilize less regulated and less specific  
522 transcription and translation mechanism. As many of these truncated transcripts are not  
523 predicted to contain canonical translation initiation sites, S1 binding could offer a  
524 mechanistic explanation for the translation of such proteins arising from unexpected  
525 transcriptional start sites.

526

527 To examine whether ribosomal protein S1-based translation initiation is likely to occur in  
528 *Prochlorococcus* MED4, we quantified the relative frequency of 10 and 12mer  
529 pyrimidine rich motifs (made up of at least 80% pyrimidines) – sequences which are  
530 conducive to S1 binding (Mutsuda and Sugiura, 2006). In connection with the  
531 pyrimidine rich regions, we also searched for NtcA binding regions based on the  
532 predicted motifs described previously (Su, 2005; Tolonen *et al.*, 2006). Sites were  
533 counted as possible NtcA binding sites if they were found less than 100 bp upstream of  
534 identified translational start sites, although in some published cases sites can be found  
535 much further upstream (Su, 2005). We found abundant 10 and 12mer pyrimidine rich  
536 regions directly upstream of annotated translational start sites (Supplementary Table  
537 S15). Furthermore, we discovered that many genes had possible NtcA binding sites,  
538 which could possibly facilitate NtcA regulation under N stress (Supplementary Table  
539 S15). There are also 912 pyrimidine rich motifs found in non-coding regions of the  
540 genome (observed:expected = 1.35). Purine rich sequences are concomitantly under-  
541 observed and, given that there are no G+C biases in purine and pyrimidine motifs, these  
542 observations are consistent with the hypothesis that S1 translation initiation can occur  
543 both in 5' untranslated regions and within sequences downstream of traditional start sites.  
544  
545 These data suggest that protein S1 could associate with these pyrimidine rich sequences  
546 and mediate translation of proteins expressed from the annotated primary start site or, as  
547 suggested by our transcriptome data under N deprivation conditions, from internally  
548 transcribed start sites. Leaderless mRNAs are likely present within *Prochlorococcus*  
549 MED4, requiring an alternate method of translation initiation. In these cases, mRNAs



550 lacking a 5' UTR directly bind 70S monosomes, thus initiating translation (Moll *et al.*,  
551 2002; Voigt *et al.*, 2014). Our data, consistent with previous studies on *Prochlorococcus*  
552 MED4 (Voigt *et al.*, 2014), suggest that 6-8% of all primary TSSs are found within 10nt  
553 of translation initiation sites and thus potentially initiated by 70S monosomes.

554

## 555 **Conclusions**

556 In summary, we have shown that a variety of structural changes occur within the  
557 *Prochlorococcus* transcriptome in response to N deprivation, and that these changes  
558 likely contribute to the ability of this organism to minimize the amount of N required in  
559 its proteome. Specifically, *Prochlorococcus* increased internal transcription for several  
560 genes during N stress conditions, which should in some cases produce shortened versions  
561 of enzymes that likely retain at least partial functionality and require fewer N atoms.  
562 Proteomic confirmation of the translation of the shortened peptides, as well as  
563 biochemical characterization, will be necessary to understand their abundance and  
564 function relative to their full-length versions. We also propose that *Prochlorococcus* may  
565 have relatively slow translation rates which, in conjunction with short RNA half-lives,  
566 allows them to control protein abundances, reducing cellular nitrogen requirements.  
567 Although the concept of cost minimization has heretofore been considered to function on  
568 evolutionary time scales (through selection on genomic codon usage to reduce nutrient  
569 requirements), these results show that cost minimization can encompass physiological  
570 mechanisms as well – through dynamic structural changes to the transcriptome that  
571 should result in a proteome requiring fewer N atoms.

572

573 **Acknowledgements**

574 We thank Alexis Yelton (MIT) and Julie Miller (MIT) for assistance with sampling. We  
575 also thank Dr. David Vuono (DRI) for his contributions in the preparation of the  
576 manuscript. This work was funded in part by the Gordon and Betty Moore Foundation  
577 through Grant GBMF495 to SWC, by a grant from the Simons Foundation (SCOPE  
578 award ID 329108 to SWC), and by the National Science Foundation (MCB-1244630 to  
579 JIG and OCE-1153588 and DBI-0424599 to SWC). ACR was supported by a HHMI  
580 International Student Research Fellowship. This article is a contribution from the Simons  
581 Collaboration on Ocean Processes and Ecology (SCOPE).  
582

583 **References:**

- 584 Ahlgren NA, Rocap G, Chisholm SW. (2006). Measurement of *Prochlorococcus*  
585 ecotypes using real-time polymerase chain reaction reveals different abundances of  
586 genotypes with similar light physiologies. *Environ Microbiol* **8**: 441–454.
- 587 Amman F, Wolfinger MT, Lorenz R, Hofacker IL, Stadler PF, Findeiß S. (2014).  
588 TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics* **15**: 89.
- 589 Anders S, Huber W. (2010). Differential expression analysis for sequence count data.  
590 *Genome Biol* **11**: R106.
- 591 Andrews S. (2009). FastQC A Quality Control tool for High Throughput Sequence Data.
- 592 Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D. (2001). Molecular evolution  
593 of protein atomic composition. *Science* **293**: 297–300.
- 594 Bibby TS, Mary I, Nield J, Partensky F, Barber J. (2003). Low-light-adapted  
595 *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**:  
596 1051–1054.
- 597 Cavender-Bares KK, Mann EL, Chisholm SW, Ondrusek ME, Bidigare RR. (1999).  
598 Differential response of equatorial Pacific phytoplankton to iron fertilization. *Limnol and*  
599 *Oceanogr* **44**: 237–246.
- 600 Celesnik H, Deana A, Belasco JG. (2007). Initiation of RNA Decay in *Escherichia coli*  
601 by 5' Pyrophosphate Removal. *Mol Cell* **27**: 79–90.
- 602 Chisholm SW. (1992). Phytoplankton Size. In: Falkowski PG, Woodhead AD (eds).  
603 *Primary Productivity and Biogeochemical Cycles in the Sea*. Springer US: Boston, MA,  
604 pp 213–237.
- 605 Deutsch C, Sarmiento JL, Sigman DM, Gruber N, Dunne JP. (2007). Spatial coupling of  
606 nitrogen inputs and losses in the ocean. *Nature* **445**: 163–167.
- 607 Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, *et al.* (2013).  
608 Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and  
609 *Synechococcus*. *Proc Natl Acad Sci U S A* **110**: 9824–9829.
- 610 Gavis J. (1976). Munk and Riley revisited - Nutrient diffusion transport and rates of  
611 phytoplankton growth. *J Mar Res* **34**: 161–179.
- 612 Gilbert JDJ, Fagan WF. (2011). Contrasting mechanisms of proteomic nitrogen thrift in  
613 *Prochlorococcus*. *Mol Ecol* **20**: 92–104.
- 614 Giovannoni SJ, Cameron Thrash J, Temperton B. (2014). Implications of streamlining  
615 theory for microbial ecology. *ISME J* **8**: 1553–1565.

- 616 Grzymalski JJ, Dussaq AM. (2012). The significance of nitrogen cost minimization in  
617 proteomes of marine microorganisms. *ISME J* **6**: 71–80.
- 618 Grzymalski JJ, Grzymalski JJ, Marsh AG, Marsh AG. (2014). Protein Languages Differ  
619 Depending on Microorganism Lifestyle Ouzounis CA (ed). *PLoS ONE* **9**: e96910 EP–.
- 620 Havaux M, Guedeney G, He Q, Grossman AR. (2003). Elimination of high-light-  
621 inducible polypeptides related to eukaryotic chlorophyll a/b-binding proteins results in  
622 aberrant photoacclimation in *Synechocystis* PCC6803. *BBA-Bioenergetics* **1557**: 21–33.
- 623 Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, *et al.* (2014). A  
624 multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion  
625 detection. *Genome Biol* **15**: R34.
- 626 Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, *et al.* (2009). Fast  
627 Mapping of Short Sequences with Mismatches, Insertions and Deletions Using Index  
628 Structures Searls DB (ed). *PLoS Comput Biol* **5**: e1000502.
- 629 Li G-W, Oh E, Weissman JS. (2012). The anti-Shine-Dalgarno sequence drives  
630 translational pausing and codon choice in bacteria. *Nature* **484**: 538–541.
- 631 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* (2009). The  
632 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- 633 Lybecker M, Bilusic I, Raghavan R. (2014). Pervasive transcription: detecting functional  
634 RNAs in bacteria. *Transcription* **5**: e944039–5.
- 635 Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, *et al.* (2010).  
636 Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans.  
637 *ISME J* **4**: 1252–1264.
- 638 McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumbly P, Genco CA, *et al.*  
639 (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res* **41**: e140–  
640 e140.
- 641 Mezulis S, Yates CM, Wass MN, Sternberg MJE, Kelley LA. (2015). The Phyre2 web  
642 portal for protein modeling, prediction and analysis. *Nature Protocols* **10**: 845–858.
- 643 Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, Bantscheff J, *et al.* (2011). An  
644 experimentally anchored map of transcriptional start sites in the model cyanobacterium  
645 *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A* **108**: 2124–2129.
- 646 Moll I, Grill S, Gualerzi CO, Bläsi U. (2002). Leaderless mRNAs in bacteria: surprises in  
647 ribosomal recruitment and translational control. *Mol Microbiol* **43**: 239–246.
- 648 Moore CM, Moore CM, Mills MM, Mills MM, Arrigo KR, Arrigo KR, *et al.* (2013).  
649 Processes and patterns of oceanic nutrient limitation. *Nature* **6**: 701–710.

- 650 Moore LR, Moore LR, Coe A, Coe A, Zinser ER, Zinser ER, *et al.* (2007). Culturing the  
651 marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr Met* **5**: 353–362.
- 652 Munk WH, Riley GA. (1952). Absorption of nutrients by aquatic plants. *J Mar Res* **11**:  
653 215–240.
- 654 Mutsuda M, Sugiura M. (2006). Translation initiation of cyanobacterial *rbcS* mRNAs  
655 requires the 38-kDa ribosomal protein S1 but not the Shine-Dalgarno sequence:  
656 development of a cyanobacterial in vitro translation system. *J Biol Chem* **281**: 38314–  
657 38321.
- 658 Ohashi Y, Shi W, Takatani N, Aichi M, Maeda S-I, Watanabe S, *et al.* (2011). Regulation  
659 of nitrate assimilation in cyanobacteria. *J Exp Bot* **62**: 1411–1424.
- 660 Olson RJ, Vaultot D, Chisholm SW. (1985). Marine phytoplankton distributions measured  
661 using shipboard flow cytometry. *Deep-Sea Res* **32**: 1273–1280.
- 662 Parkhill JP, Maillet G, Cullen JJ. (2001). Fluorescence-based maximal quantum yield  
663 for PSII as a diagnostic of nutrient stress. *J Phycol* **37**: 517–529.
- 664 Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, *et al.*  
665 (2011). Integrative genomics viewer. *Nat Biotech* **29**: 24–26.
- 666 Sauer J, Görl M, Forchhammer K. (1999). Nitrogen starvation in *Synechococcus* PCC  
667 7942: involvement of glutamine synthetase and NtcA in phycobiliprotein degradation and  
668 survival. *Arch Microbiol* **172**: 247–255.
- 669 Schoenberg DR. (2007). The end defines the means in bacterial mRNA decay. *Nat Chem*  
670 *Biol* **3**: 535–536.
- 671 Shao W, Price MN, Deutschbauer AM, Romine MF, Arkin AP. (2014). Conservation of  
672 Transcription Start Sites within Genes across a Bacterial Genus. *mBio* **5**: e01398–14–  
673 e01398–14.
- 674 Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiß S, Sittka A, *et al.* (2010). The  
675 primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**:  
676 250–255.
- 677 Sherman MY, Qian S-B. (2013). Less is more: improving proteostasis by translation slow  
678 down. *Trends Biochem Sci* **38**: 585–591.
- 679 Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, *et al.*  
680 (2008). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes  
681 in the Sargasso Sea. *ISME J* **3**: 93–105.
- 682 Steglich C, Futschik ME, Futschik ME, Lindell D, Lindell D, Voss B, *et al.* (2008). The  
683 Challenge of Regulation in a Minimal Photoautotroph: Non-Coding RNAs in  
684 *Prochlorococcus* Matic I (ed). *PLoS Genet* **4**: e1000173.

- 685 Steglich C, Lindell D, Futschik M, Rector T, Steen R, Chisholm SW. (2010). Short RNA  
686 half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biol*  
687 **11**: R54.
- 688 Su Z. (2005). Comparative genomics analysis of NtcA regulons in cyanobacteria:  
689 regulation of nitrogen assimilation and its coupling to photosynthesis. *Nucleic Acids Res*  
690 **33**: 5156–5171.
- 691 Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, *et al.*  
692 (2015). Global transcriptional start site mapping using differential RNA sequencing  
693 reveals novel antisense RNAs in *Escherichia coli*. *J Bacteriol* **197**: 18–28.
- 694 Thompson AW, Huang K, Saito MA, Chisholm SW. (2011). Transcriptome response of  
695 high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME*  
696 *J* **5**: 1580–1594.
- 697 Thorvaldsdóttir H, Robinson JT, Mesirov JP. (2013). Integrative Genomics Viewer  
698 (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*  
699 **14**: 178–192.
- 700 Tjaden B. (2015). De novo assembly of bacterial transcriptomes from RNA-seq data.  
701 *Genome Biol* **16**: 1.
- 702 Tolonen AC, Tolonen AC, Aach J, Aach J, Lindell D, Lindell D, *et al.* (2006). Global  
703 gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen  
704 availability. *Mol Syst Biol* **2**: 53.
- 705 Tyrrell T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary  
706 production. *Nature* **400**: 525–531.
- 707 Van Mooy BAS, Rocap G, Fredricks HF, Evans CT, Devol AH. (2006). Sulfolipids  
708 dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine  
709 environments. *Proc Natl Acad Sci U S A* **103**: 8607–8612.
- 710 Voigt K, Sharma CM, Mitschke J, Lambrecht SJ, Voß B, Hess WR, *et al.* (2014).  
711 Comparative transcriptomics of two environmentally relevant cyanobacteria reveals  
712 unexpected transcriptome diversity. *ISME J* **8**: 2056–2068.
- 713 Wickham H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.
- 714 Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ, *et al.* (2006).  
715 *Prochlorococcus* Ecotype Abundances in the North Atlantic Ocean As Revealed by an  
716 Improved Quantitative PCR Method. *Appl Environ Microbiol* **72**: 723–732.
- 717 Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW. (2007). Influence  
718 of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean.  
719 *Limnol and Oceanogr* **52**: 2205–2220.

720

721 **Figure Legends:**

722 **Figure 1. Physiological changes of *Prochlorococcus* MED4 during N deprivation. A.**

723 Effect of N deprivation on cell concentration. **B.** Effect of N deprivation on bulk culture  
724 fluorescence. **C.** Effect of N deprivation on maximum quantum efficiency of  
725 photosystem II (Fv/Fm) as measured by fast repetition rate fluorometry. Black lines  
726 represent the original triplicate cultures grown in N-replete media; other lines indicate  
727 cultures pelleted and resuspended in either N-replete (dark grey line) or N-deficient (light  
728 grey line) media. The discontinuity in fluorescence and cell concentration measurements  
729 result from incomplete recovery of cells following centrifugation. Error bars represent the  
730 standard deviation of 3 biological replicates and are smaller than the symbols when not  
731 visible.

732

733 **Figure 2. Primary transcriptome mapping to 3 genes during N deprivation.** Values  
734 indicate the number of primary reads mapping to the **A.** PMM0149 (*ndhF*), **B.** PMM0058  
735 (conserved hypothetical protein) and **C.** PMM1485 (*rpoB*) genes. Experimental (red) and  
736 control reads (black) were compared at 12 hours post N deprivation. The primary and  
737 internal TSSs are marked with arrows in each panel. Each panel represents the full length  
738 of the gene, with the x-axis representing the distance from the primary start site.

739

740 **Figure 3. Structure predictions for internal transcription sites versus PDB**

741 **structures. A.** Protein structure threading for the cyanate ABC transporter (PDB  
742 c2i4cA). There are no major structural differences in the protein threads. Blue and red are  
743 the overlapping structure representing the full length transcript (blue) and the

744 corresponding predicted protein from the internal start site in MED4 (red). **B.** As in A,  
745 but for the Ferredoxin-NADP reductase protein (PDB C1jb9A). **C.** Protein structure  
746 threading for the for RNA polymerase (PDB C3lu0C).

747

748 **Figure 4. Frequency of Glycine-Glycine motifs in selected microbial genomes.**

749 Deviations in observed to expected ratios of Gly-Gly motifs, indicative of the potential  
750 for ribosomal stalling, are indicated for a set of common copiotrophic r-selected and  
751 oligotrophic k-selected organisms.

752



Table 1. Differentially Expressed Transcripts between N-deprived vs. N-replete Conditions Per Timepoint ( $p < 0.05$ )

Time (hr)	Regulation	Amount of Genes	Differentially
		Differentially Expressed Genes	Expressed (%)
3	Up	62	7.21%
12	Up	84	9.77%
24	Up	93	10.81%
3	Down	28	3.26%
12	Down	86	10.00%
24	Down	107	12.44%

Table 2. Transcriptional Start Sites By Category

Type	Downstream Antisense	Internal Antisense	Internal	Internal Antisense Downstream	Internal Primary	Orphan	Primary	Primary Antisense
N-Replete	6	363	502	1	122	124	542	26
N-Starved	8	275	519	5	116	118	477	20

## Figure 1

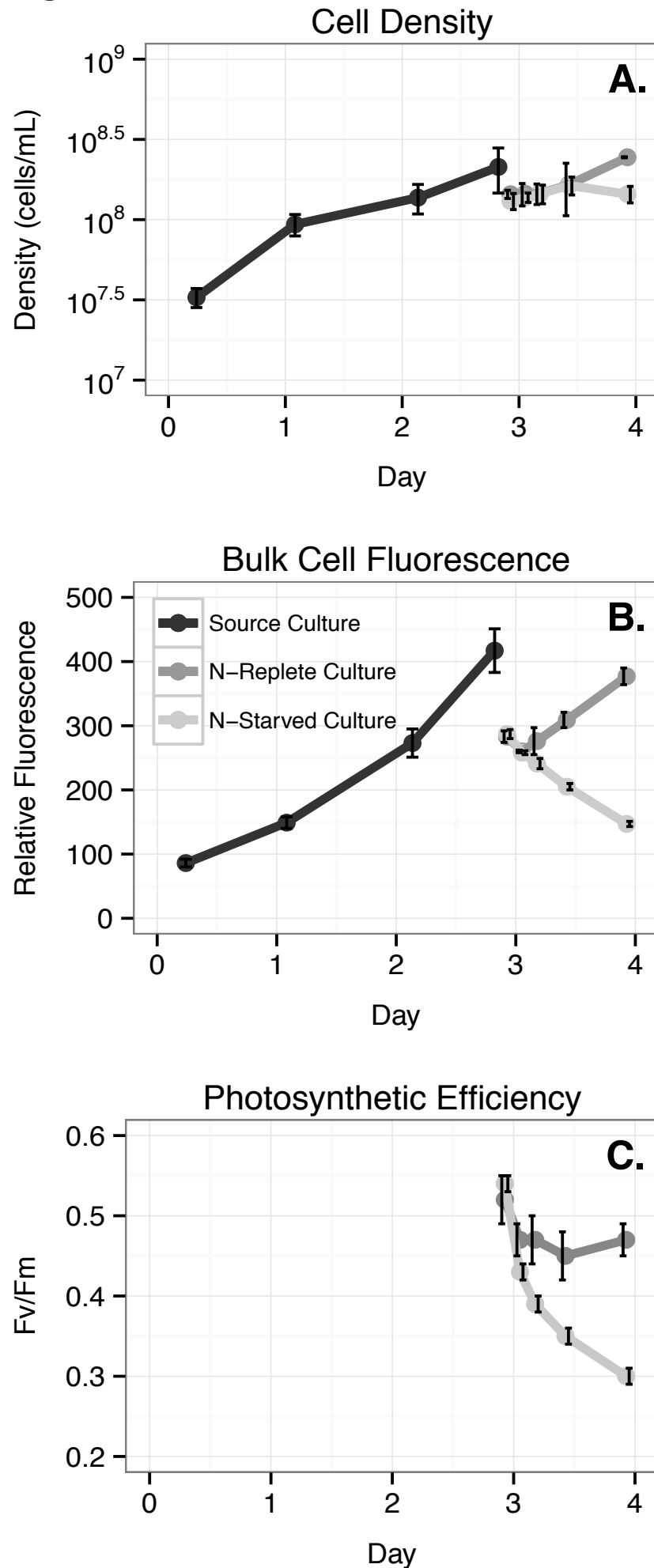


Figure 2

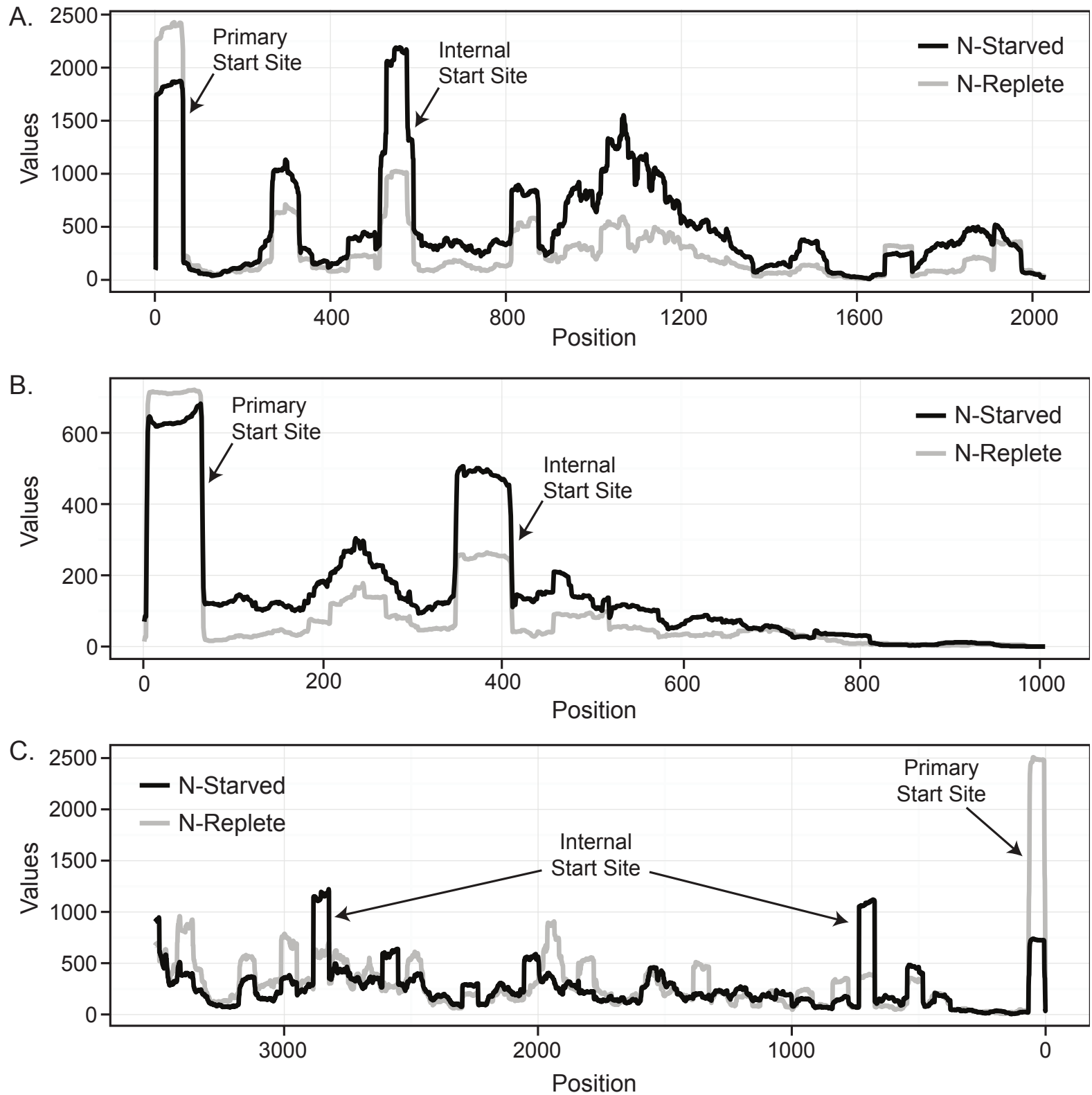


Figure 3

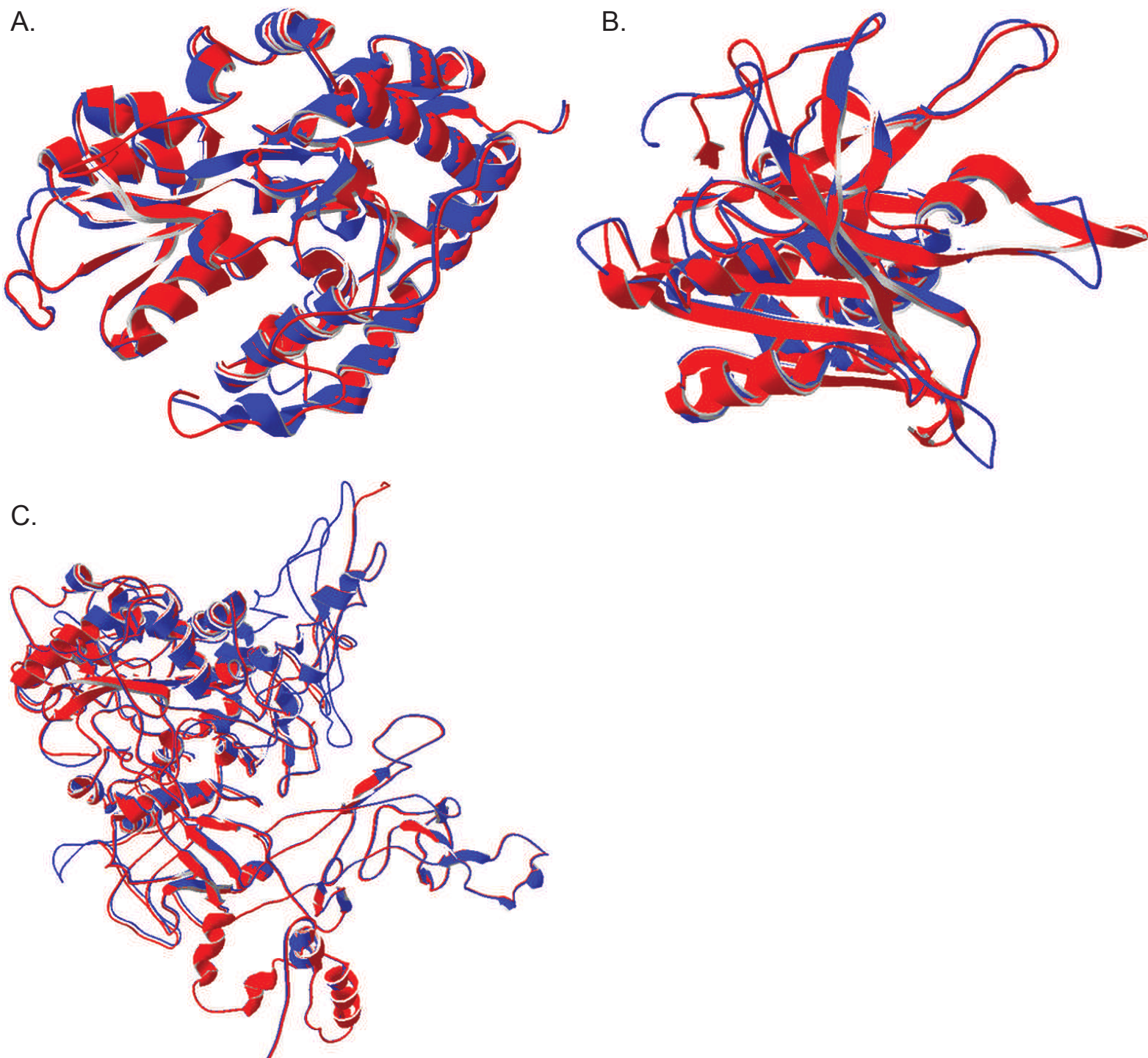


Figure 4

